Check for
updates

Farm-Level Smart Crop Recommendation Framework Using Machine Learning

Amit Bhola[1] · Prabhat Kumar[1]

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

Abstract

Agriculture is the primary source of food, fuel, and raw materials and is vital to any country's economy. Farmers, the backbone of agriculture, primarily rely on instinct to determine what crops to plant in any given season. They are comfortable following customary farming practices and standards and are oblivious to the fact that crop yield is highly dependent on current environmental and soil conditions. Crop recommendations involve multifaceted factors such as weather, soil quality, crop production, market demand, and prices, making it crucial for farmers to make well-informed decisions. An improper or imprudent crop recommendation can affect them, their families, and the entire agricultural sector. Modern technologies like artificial intelligence, machine learning, and data science have emerged as efficient solutions to combat issues like declining crop production and lower profits. This research proposes a Smart Crop Recommendation framework that leverages machine learning to empower farmers to make informed decisions about optimal crop selection. The framework consists of two phases: crop filtration and yield prediction. Crops are filtered in the first phase using an artificial neural network based on local input parameters. The second phase estimates yield for filtered crops, considering the season, farm area, and location data. The final recommendation provides farmers with crops aimed at maximizing profit. The remarkable 99.10% accuracy of the framework is demonstrated through experimentation using artificial neural networks and the 0.99 R^2 error metric for the random forest. The uniqueness of this framework lies in its distinctive focus on the farm level and its consideration of the challenges and various agricultural features that change over time. The experimental results affirm the effectiveness of the framework, and its lightweight nature enhances its practicality, making it an efficient real-time recommendation solution.

✉ Amit Bhola
amitb.phd19.cs@nitp.ac.in

Prabhat Kumar
prabhat@nitp.ac.in

[1] Department of Computer Science and Engineering, National Institute of Technology Patna, Patna, Bihar, India

🖉 Springer

1 Introduction

The agriculture sector is a crucial source of livelihood for people all around the globe. In addition to being the primary source of food, this sector plays a pivotal role in a country's economy and produces employment opportunities. Worldwide, India ranks second among food producers with more than a 50% contribution towards total national employment [1]. Furthermore, the share of agriculture and allied sectors for 2022–2023 reached 18.3% of the country's gross value added [2]. However, rapid population growth has increased food demand, thus putting pressure on agricultural productivity. Moreover, 86% of all farmers in India are small and marginal, holding less than two hectares of land [3].

Furthermore, these small and marginal farmers still use conventional or customary farming practices. For instance, they use their primitive knowledge for crop selection by preferring traditional or popular crops in their region. As a result, crop yield and land fertility may suffer impacting farming returns. Increased soil acidity is one of the major fallouts of inappropriate crop selection and insufficient soil nutrients. Further, environmental conditions, climate variability, and water levels influence crop quality and productivity. Crop selection is an essential factor in increasing agricultural productivity and quality. Motivated by the above challenges, this research identifies and addresses issues that arise in the production of crops.

The fusion of contemporary technologies with conventional sectors also has the capacity to revolutionize established practices, and a notable instance of this synergy is the incorporation of data science and data mining in agriculture [4]. Data science in agriculture is emerging as a transformative power, providing sophisticated analytics and predictive modeling methods that assist farmers in maximizing yields, managing resources, and adapting to shifting environmental conditions. The integration of data science techniques with agricultural practices is reshaping farming methods for the modern era, leading to increased productivity, and reduced costs. Further, the utilization of cutting-edge technologies like machine learning (ML) [5], deep learning [6, 7], big data analytics [8], and the internet of things [9, 10] have proven to be highly beneficial to the agricultural industry [11].

A machine learning based Smart Crop Recommendation (SCR) framework is proposed to address the crop selection dilemma by considering factors such as temperature, precipitation, soil pH, humidity, and soil nutrients. In-situ soil and environmental factors are important for farming. Specifically, soil nutrients such as Nitrogen (N), Phosphorus (P), and Potassium (K) are essential for plant growth and preventing disease. In contrast, soil pH controls the chemical reactions by checking its acidic or alkaline level. In addition, the development of plants is highly affected by soil electical conductivity, indicating soil fertility, water quality, and salinity. Rainfall is another critical factor for crop yield, as different crops may require different amounts of water.

The government has made efforts to improve agricultural productivity by providing soil health cards (SHCs) to individual farmers after analyzing the soil of their farms.

SHC contains macro and micro soil nutrient levels corresponding to the farm. However, farmers' traditional cultivation approaches fail to utilize this information to improve agricultural productivity.

The SCR framework is simple and cost-effective that uses machine learning techniques to recommend crops based on local parameters. The main contributions of this paper are summarized as follows:

- A Smart Crop Recommendation framework comprising two distinct phases is proposed to guide farmers in selecting optimal crops for enhanced returns.
- The initial phase incorporates an artificial neural network (ANN) model designed to filter out unsuitable crops by considering farm soil nutrients and regional weather conditions.
- The subsequent phase employs a regressor model utilizing the random forest to predict crop yields accurately.
- The final crop recommendations are determined based on maximizing profit, considering both the cost of production (COP) and market price (MP).
- Extensive experiments are conducted to demonstrate the efficacy of the proposed framework.

The rest of the paper is structured as follows: Sect. 2 provides a summary of related works, and Sect. 3 delves into the proposed methodology for the crop recommendation framework. In Sect. 4, the experimental setup, data exploration, and evaluation metrics are detailed. The analysis of the results for both phases is presented in Sect. 5, followed by a discussion. Finally, Sect. 6 concludes the work with the scope for future research.

2 Related Works

Digital agriculture, also known as smart farming or e-agriculture, is the utilization of digital tools and technologies to collect, store, analyze, share and mine electonic data within the agricultural sector [12, 13]. Research in digital agriculture has made significant progress toward its goals of crop selection, yield prediction and real-time farm management [14]. Cheema et al. [15] devised a diversified crop model utilizing various soil parameters to identify suitable crops. Their model employed a quantum value-based gravitational search algorithm (GSA) to optimize solutions, considering soil factors like pH, salinity, texture, nitrogen, phosphorous, and potassium as inputs for crop selection. Bakthavatchalam et al. [16] proposed a crop prediction system leveraging multilayer perceptron, JRip, and decision table classifiers based on diverse attributes. WEKA tool implementation showcased that multilayer perceptron (MLP) was the best-performing model with an impressive accuracy of 98.22%.

Jain et al. [17] proposed a soil-based machine learning comparative analytical framework that assesses soil characteristics and climate factors to predict crop yield classes (high, low, and medium). The result for comparative analysis demonstrates that support vector machine (SVM) achieved maximum accuracies of 85.62% and 75.64% for wheat and maize, respectively. Gupta et al. [18] presented a crop recommendation system integrating MapReduce and K-means clustering, considering crop yields per acre for various regions and different varieties grown in the target area. Mariammal et

Table 1 Summary of the relevant works

Reference	Year	Type	Technique	Dataset	Analysis
Cheema et al. [15]	2021	CS	GSA	Soil parameters	Crop preference is generated w.r.t. soil properties and top crop is selected
Bakthavatchalam et al. [16]	2022	CS	MLP	Rainfall, temp., soil	WEKA based crop predictions for 22 crops
Jain et al. [17]	2022	CYP	SVM, Naive Bayes, RF, LR	Soil and climate parameters crop: wheat, maize	Recommended crop yield using SMLF framework
Gupta et al. [18]	2021	CS	MapReduce framework, K-means	Rainfall, humidity, temp., soil	A self-designed recommender system has been used to predict the crops
Mariammal et al. [19]	2021	CR	RF, NB, DT, kNN, SVM, FS: MRFE	Soil and environmental parameters	MRFE outperform other FS with an Acc. of 95% in crop prediction
Khosla et al. [20]	2020	CYP	RNN (rainfall prediction), SVR (CYP)	Rainfall	Predicted yield for four kharif crops in Visakhapatnam
Gopal et al. [21]	2019	CYP	MLR-ANN, SVR, kNN, RF	Area, tanks, open wells, temp	Hybrid MLR-ANN outperforms other models in CYP with R value of 0.99
Elavarasan et al. [22]	2020	CYP	RNN, Q-learning network	Climate and soil parameters crop: paddy	Proposed reinforcement model that predicts yield
Daniel et al. [23]	2022	CR	Deep neural network	Soil parameters	Suggestion of suitable crops and cost profits
Iniyan et al. [24]	2022	CYP	RF, KNN, DT, MLR, MSE	Precipitation, solar radiation, temperature crop: corn and soybean	Proposed MIFS based multilevel stacked ensemble regression

Table 1 continued

Reference	Year	Type	Technique	Dataset	Analysis
Kuradusenge et al. [25]	2023	CYP	RF, SVR, PR	Rainfall, temp. crop: potato, maize	Crop yield prediction for Irish potato and maize
Swathi et al. [26]	2023	CS	DT, SVM, NB, KNN, XGB, RF	Soil parameters	Suitable crop based on the soil nutrition levels is proposed
Bandi et al. [27]	2023	CR	Ensemble ML model	Soil parameters	Proposed a crop recommendation system to address the issue of increasing crop production
Shingade et al. [28]	2023	CR	Soil, weather	DT, SVM, RF	Crop recommendation using the sensor information collected by 250 sensors located in different Maharashtra places
Shams et al. [29]	2024	CYP	GB, DT, RF, GNB, MNB	Season, area, production	Proposed XAI-CROP, a crop recommendation system leveraging explainable artificial intelligence (XAI) using Yield prediction
Olisah et al. [30]	2024	CYP	DNNR, RF, XGB	Soil and climate parameters	Proposed corn yield prediction model for smallholder farmer
Proposed work	–	CR, CYP	ANN, RF	Soil, weather, season, area, production, COP, MP	Crop recommendation framework comprising two phase to guide farmers in selecting optimal crops for enhanced returns

FS, feature selection; CS, crop selection; CR, crop recommendation; CYP, crop yield prediction; GSA, gravitational search algorithm; Acc., accuracy; RF, random forest; NB, Naïve Bayes; MRFE, modified recursive feature elimination; MANN, modular ANN; MLR, multiple linear regression; SVR, support vector regression; PR, polynomial regression; RNN, recurrent neural network; GB, gradient boosting; DT, decision tree; DNNR, deep neural network regressor; XAI, explainable artificial intelligence; COP, cost of production; MP, market price

al. [19] proposed a feature selection technique named modified recursive feature elimination (MRFE) for crop prediction, aiming to identify essential features from crop data. Their approach demonstrated that MRFE outperformed various wrapper-based feature selection techniques utilizing a ranking algorithm, achieving an accuracy of 95%. Shams et al. [29] proposed XAI-CROP, a crop recommendation system leveraging explainable artificial intelligence (XAI) for transparency. The study extensively compares XAI-CROP with various machine learning models, demonstrating superior performance through low MSE (0.9412) and MAE (0.9874), indicating highly accurate crop yield predictions. The robust R^2 value of 0.94152 emphasizes XAI-CROP's ability to explain 94.15% of data variability, showcasing its interpretability and reliability.

Swathi et al. [26] proposed a model for crop classification and prediction based on soil nutrition in India to address issues of low yield. Various machine learning models are employed on datasets collected from Kaggle, including six crop types and 11 nutrients. The results indicate that extreme gradient boosting and naive bayes outperform other models with AUC scores of 0.994 and 0.993, respectively. Bandi et al. [27] proposed a voting classifier-based crop recommendation system, leveraging machine learning to enhance precision agriculture. The system addresses the challenges faced by farmers in optimizing crop production based on climate and soil properties by utilizing ensemble modeling with majority voting, and it achieved an impressive accuracy of 99.4%. This approach aims to minimize financial losses for farmers and enhance informed agricultural decision-making.

Khosla et al. [20] employed various models, such as support vector regression (SVR), random forest (RF), linear regression, and k-nearest neighbors, to predict crop yield across four major Kharif crops. They initially forecasted rainfall using a modular artificial neural network model and utilized this prediction as input to SVR for crop yield estimation, revealing that SVR outperformed among various machine learning models. Gopal et al. [21] introduced a hybrid model combining multiple linear regression (MLR) and an ANN for yield forecasting. The hybrid model utilized MLR intercept and coefficients to initialize input layer weights and bias, showcasing superior performance with paddy crops compared to conventional ML techniques. Devi et al. [31] conducted a study to identify significant factors affecting agricultural production, utilizing ordinary least squares and ridge regression. Time series data were used to measure the variability in the area, production, and yield for four selected crops based on adjusted R^2 and RMSE values.

Elavarasan et al. [22] leveraged reinforcement learning to predict crop yield. They employed a Deep Recurrent Q-Network based on input parameters, achieving a notable 93.7% accuracy in crop yield anticipation. Olisah et al. [30] presented a deep neural network regressor (DNNR) for corn yield prediction to address the interaction between weather and soil variables. Outperforming random forest and extreme gradient boosting regressors, the DNNR achieved impressively low prediction errors of 0.0146 t/ha and 0.0209 t/ha. The study emphasized empowering smallholder farmers with a mobile application decision support system, thus incorporating education and farmer-to-market access modules for intelligent farming decisions and potential impact on food crises. Daniel et al. [23] proposed a web application for crop recommendation to aid farmers in selecting effective crops and organic fertilizers. Their

algorithm incorporated a deep neural network to predict prices, enhancing farmers' decision-making in crop selection.

Table 1 provides a comparative analysis of various studies on crop recommendation and yield prediction. The analysis reveals a significant need for improvement in the literature, as existing works have focused on a limited set of parameters to predict crop suitability. For example, [15, 17], and [23] solely utilized soil parameters to recommend crops, while [20] relied exclusively on rainfall as an input parameter for predicting crop yield. Notably, a specific soil type may be suitable for various crops, but the yield can be adversely affected if climatic conditions are unfavorable. This research proposes a crop prediction architecture to address these limitations.

3 Methodology

Small and marginal farmers frequently find themselves entangled in the cycle of decreased production, leading to insufficient earnings, limited savings, and minimal investments. Their struggles with reduced crop yields and profits are rooted in the need to comprehend crop selection and the factors influencing their growth. Since crop selection is the most critical factor in maximizing crop yield and profitability, this work aims to develop a smart crop recommendation framework for enhancing agricultural returns. The framework helps farmers decide on suitable crops based on various local parameters.

Let, $C_1, C_2, ..., C_i$ be 'i' different crops and $F_1, F_2, ..., F_j$ be 'j' different farmlands. Each farmer is assumed to have a SHC for their farmland 'F_j' providing the soil nutrition level and regular metrological updates by government agencies. The goal is to find suitable crops for each farmland based on soil and weather inputs. Figure 1 presents the proposed framework that recommends diverse crops for each farm using a two-phase process. The first phase filters 'n' crops for each land 'F_j' from the available 'i' crops. This phase matches the suitability of various crops corresponding to the local soil and weather conditions. The filtered crops are then fed to the next step for further analysis. The second phase estimates the yield for each crop on the available farmer's land. The yield estimation helps compute the return for each crop individually. Further, the cost of cultivation and market price are used to estimate the net profit for each crop and recommend a list of crops to farmers along with the net profit. Each of these phases is further elaborated in the following subsections.

3.1 Crop Filter

Figure 2 depicts the first phase that filters 'n' suitable crops. For each farm 'F_j', let $w_1(t), w_2(t), ..., w_k(t)$ be the weather conditions, such as temperature, rainfall at time t, and let $s_1(t), s_2(t), ..., s_l(t)$ be the soil attributes such as N, P, K. The regular weather update $w_k(t)$ at time 't' is provided to the farmers by the metrological department or local government agencies for advance planning. In addition, the government provided soil health card contains 12 essential soil macro- and micro-nutrients, including pH, electrical conductivity (EC), Organic Carbon (OC), nitrogen (N), phosphorus (P),
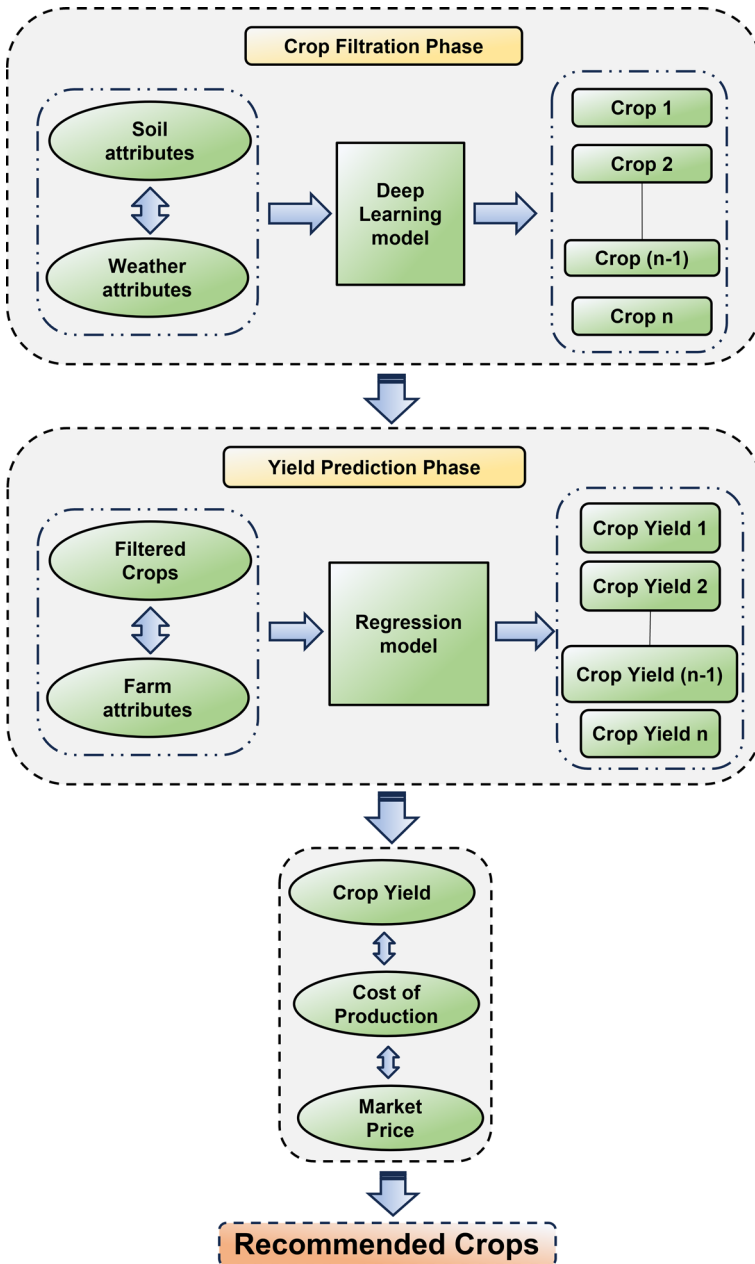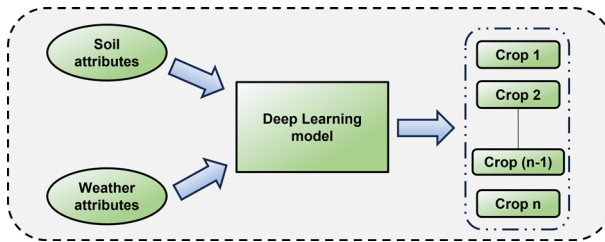
Fig. 1 Proposed framework

Fig. 2 Crop filtration phase

potassium (K), sulphur (S), zinc (Zn), boron (B), iron (Fe), Manganese (Mn), Copper (Cu). Crop growth is directly impacted by weather and soil conditions. Hence, these important soil parameters are fetched from the farmer's SHC and weather updates from government agencies to filter out the most suitable crops. A deep learning model is proposed to compute probabilities p_1, p_2,...., p_x, using input parameters and rank crops based on probabilities. Finally, the top 'n' crops are filtered and passed on to the second phase for further analysis.

Figure 3 depicts the architecture of the proposed ANN model used for the first phase. In an ANN feed-forward backpropagation network, the weights and bias, number of hidden layers, hidden neurons, learning rate and number of training epochs are essential parameters affecting prediction accuracy. Hence, the trial and error method has been used to select the parameter values for accurate predictions. A total of seven inputs are given to the input layer, and a weighted sum of inputs and bias is given as input to the hidden layer. ReLu is used as the activation function for the hidden layers, whereas the softmax activation function is used as the output layer to predict probabilities. Further, the hidden layer contains 512 nodes each, whereas the input layer contains seven nodes, and the output layer contains 17 nodes for each crop.

3.2 Crop Yield Prediction

The second phase of the framework deals with the yield prediction for each of the 'n' filtered crops obtained from the first phase. Let L_1, L_2,...., L_p be the farm location and A_1, A_2,...., A_q be the farm area for the target lands. Figure 4 depicts the second phase of the SCR framework that predicts yield using a regression model for each of the filtered crops individually on the available farmer's land L_p with an area of A_q. The model takes input filtered crops, season, location, and farm size and predicts yield for the land. Different regression models such as multiple linear regression, random forest regression, support vector regression (SVR), and XGBoost regression are applied to identify the best-performing regressor for the proposed framework.

4 Experiment

This section empirically evaluates the performance of the proposed architecture. It begins by discussing the experimental setup, the dataset source, and the data analysis.

Fig. 3 Proposed ANN
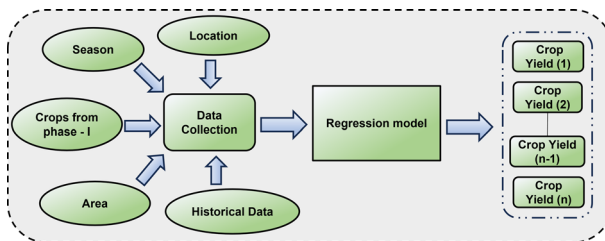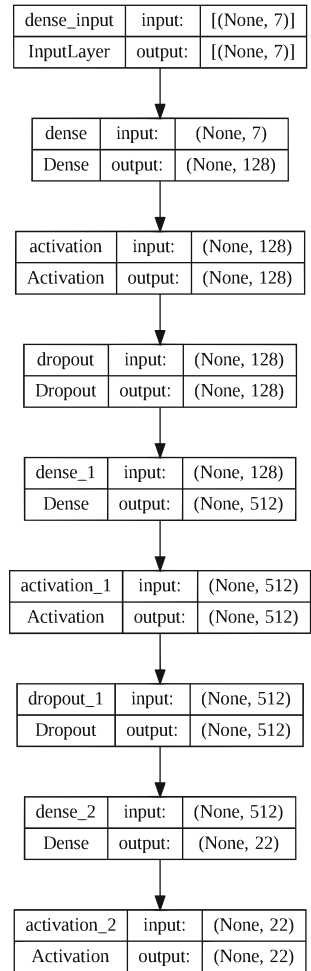architecture for crop filtration

dense_input	input:	[(None, 7)]
InputLayer	output:	[(None, 7)]

dense	input:	(None, 7)
Dense	output:	(None, 128)

activation	input:	(None, 128)
Activation	output:	(None, 128)

dropout	input:	(None, 128)
Dropout	output:	(None, 128)

dense_1	input:	(None, 128)
Dense	output:	(None, 512)

activation_1	input:	(None, 512)
Activation	output:	(None, 512)

dropout_1	input:	(None, 512)
Dropout	output:	(None, 512)

dense_2	input:	(None, 512)
Dense	output:	(None, 22)

activation_2	input:	(None, 22)
Activation	output:	(None, 22)

Fig. 4 Crop yield prediction phase

Table 2 Feature description (crop filtration phase)

Features	Description	Unit
Nitrogen (N)	It is responsible for photosynthesis in the plant	kg/ha
Phosphorus (P)	It is crucial to the crop's development	kg/ha
Potassium (K)	It is required for the reproduction of crops	kg/ha
pH level (pH)	It determines the availability of essential plant nutrients	pH value
Temperature	It is a key factor in plant growth and development	Degree celsius
Humidity	It is important for photosynthesis in plants	%
Rainfall	It is the primary source of water for agricultural production	mm

Table 3 Feature description (yield prediction phase)

Features	Description/Unit
State name	Indian states
District name	District of selected Indian state
Crop	17
Season	Kharif, Rabi, whole year
Area	Cultivated territory (hectares)
Production	Tonne
Yield	Tonne/hectare

Following this, the implementation details of the proposed model are elaborated, along with the algorithmic steps and the employed evaluation metrics.

4.1 Experimental Setup

The experimental configuration includes an Intel Core i5 processor with a 3.6 GHz and 8 GB RAM. Python served as the programming language and Google Colab Notebook was used for the program execution. Standard software libraries, including Keras, TensorFlow, Matplotlib, and NumPy, were employed.

4.2 Dataset

Two different datasets were used to evaluate the performance of the proposed SCR framework. The crop filtration dataset is obtained from Kaggle [32]. In this dataset, lands and crops are classified based on several weather and soil properties, including 2200 land and 22 crop samples. However, only 17 crops, maize, rice,...., and pomegranates, are considered for this phase due to data availability for these crops in the next phase. Table 2 describes the features used in the first phase of the framework.

The dataset utilized in the second phase was sourced from the Department of Economics and Statistics, Government of India [33]. Although the dataset comprises more than 30 crops, the selection was narrowed to 17 crops common to data availability for
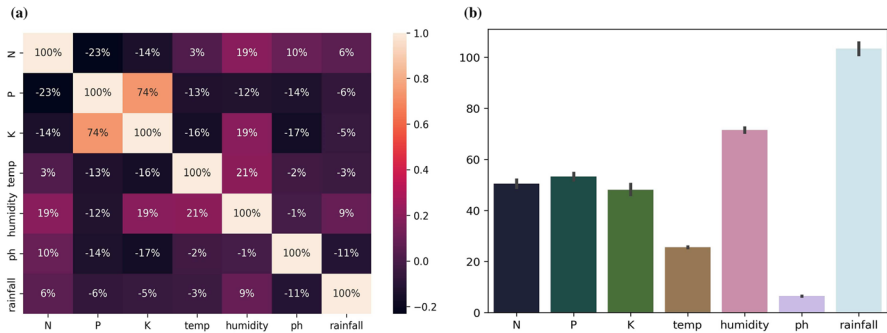
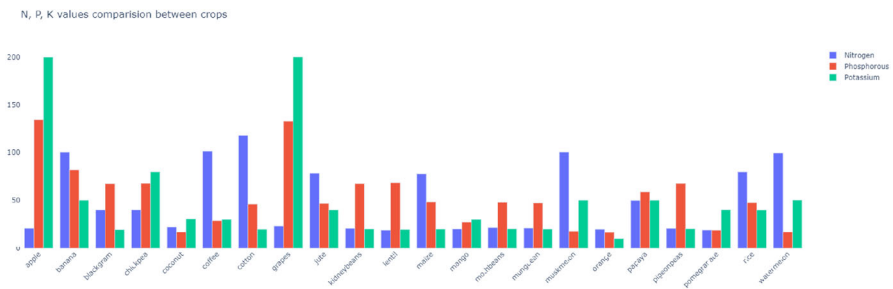**Fig. 5** Phase I dataset (**a**) correlation matrix (**b**) and feature importance

Fig. 6 N, P, and K values required by different crops

both the phases. Table 3 presents the attributes of the collected dataset for the crop yield analysis.

4.3 Dataset Analysis

This section examines the soil and environmental data affecting crop filtration and yield prediction procedures. Macronutrients like N, P, and K are substantially required for proper crop development. Figure 5a illustrates the correlation among the utilized features, emphasizing a high correlation between potassium and phosphorous as soil parameters, while humidity and rainfall show a moderate correlation. Figure 5b identifies the pivotal features in the crop filtration dataset, underscoring the significance of rain and humidity.

Figure 6 illustrates the comparison of nitrogen (N), phosphorus (P), and potassium (K) values needed by different crops. Cotton, apples, and grapes exhibit the highest macronutrient requirements for optimal growth, while lentils, black gram, and oranges have the lowest. The significance of soil macronutrients such as N, P, and K is relatively consistent across all crops. Overall, rainfall emerges as the most critical factor, with pH being the least influential among the specified parameters. Figure 7 presents the correlation matrix for the features used in the crop yield dataset.

Fig. 7 Correlation matrix (crop yield data)
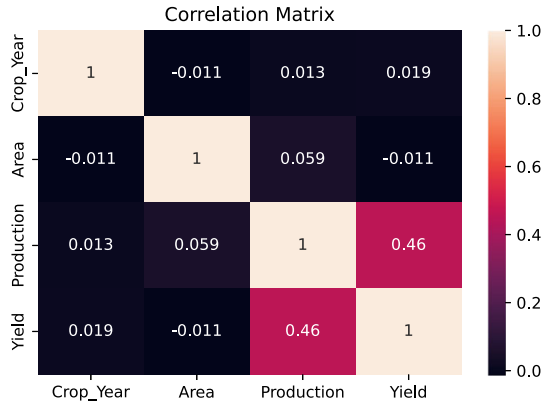
4.4 Algorithm for the SCR Framework

The primary goal of this experiment is to develop a recommendation system that will advise farmers on which crops to plant based on various factors such as soil constituents, crop traits, and climate. Algorithm 1 presents with the detailed steps involved in crop selection using ANN-RF.

Algorithm 1 Top m Crop Recommendation

1: **Input:** Soil and environment dataset **D**, Hyperparameters H
2: Pre-process the data:
3:   Normalize: $\mathbf{D} \rightarrow \mathbf{D}_{\text{norm}}$
4:   Handle missing values: $\mathbf{D}_{\text{norm}} \rightarrow \mathbf{D}_{\text{clean}}$
5: Split data into training and testing sets: $\mathbf{D}_{\text{clean}} = \{\mathbf{D}_{\text{train}}, \mathbf{D}_{\text{test}}\}$ with ratio 80:20
6: Design ANN model with hyperparameters:
7:   Hidden layers: H_{layers}
8:   Epochs: H_{epochs}
9:   Learning rate: $H_{\text{learning\_rate}}$
10: **Train Model:**
11: Initialize weights **W** and biases **b** randomly
12: **for** each epoch $e = 1$ to H_{epochs} **do**
13:   **for** each hidden layer $i = 1$ to H_{layers} **do**
14:     Compute weighted sum: $z_i = \mathbf{W}_i \cdot \mathbf{x} + \mathbf{b}_i$
15:     Apply activation: $a_i = \sigma(z_i)$
16:   **end for**
17:   Compute output layer: $z_{\text{out}} = \mathbf{W}_{\text{out}} \cdot \mathbf{a}_{H_{\text{layers}}} + \mathbf{b}_{\text{out}}$
18:   Apply softmax: $\hat{y} = \text{softmax}(z_{\text{out}})$
19: **end for**
20: Filter top n crops based on probabilities: $\{\text{Crop}_1, \ldots, \text{Crop}_n\}$
21: **Input:** Top n crops, season s, location l, area A
22: Predict yield Y for n crops using Random Forest:
23:   $Y_j = \text{RF}(\text{Crop}_j, s, l, A)$ for $j = 1$ to n
24: Calculate profit P for each crop:
25:   $P_j = Y_j \times \text{Price}_j - \text{Cost}_j$
26: Recommend top m crops based on maximum profit: $\{\text{Crop}_1, \ldots, \text{Crop}_m\}$

The algorithm is divided into two parts. The first part computes each crop's rank and filters the top 'n' crops, and the second part predicts the yield for each crop corresponding to the farmer's land. It requires soil health card details and environmental values concerning each land as input. Further, the top 'm' crops are recommended based on the maximum profit for each crop individually.

4.5 Evaluation Metrics

For multiclass recommendation tasks, accuracy serves as the most straightforward metric when considering only instances for which the predicted and true categories match. The accuracy is defined in Eq. 1, where TP (true positive) indicates instances correctly predicted as positive, FN (false negative) represents instances incorrectly predicted as negative, FP (false positive) denotes instances incorrectly predicted as positive, and TN (True Negative) signifies instances correctly predicted as negative. Additionally, for a regression evaluation, error metrics like mean absolute error (MAE), root mean square error (RMSE), and coefficient of determination (R^2) are employed to assess the model's performance.

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \tag{1}$$

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |(Y_i - \hat{Y}_i)| \tag{2}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2} \tag{3}$$

$$R^2 = 1 - \frac{\sum (Y_i - \hat{Y}_i)^2}{\sum (Y_i - \bar{Y}_i)^2} \tag{4}$$

Here, Y_i is the actual value, $\hat{Y}_i$ is the predicted value, and $\bar{Y}_i$ is the mean value. Mean absolute error (MAE) in Eq. 2 calculates the average absolute variance between predicted and actual values, offering robustness to outliers by treating all errors linearly. Root mean square error (RMSE) in Eq. 3 measures the Euclidean distance between predictions and actual values, providing an advantage in representing errors in the same unit as predicted. It assigns greater weights to larger errors. R^2 in Eq. 4 is a statistical metric indicating the goodness of fit for a regression model, with values ranging from 0 to 1.

5 Results and Discussion

In this section, the effectiveness of the proposed crop recommendation framework has been rigorously evaluated through comparative experiments. The results of these experiments are comprehensively analyzed and compared with the results produced by state-of-the-art approaches. Additionally, the adaptability of the framework is explored

Table 4 Comparative analysis (phase I)

Accuracy/models	DT	SVM	RF	KNN	XGBoost	ANN
Training	98.50	98.58	99.20	98.82	99.20	**99.27**
Testing	97.41	97.64	98.58	97.64	98.11	**99.10**

Table 5 List of filtered crops for sample land

Land	Top filtered crops
Land 1	Watermelon, papaya, grapes, chickpea, apple
Land 2	Chickpea, mothbeans, muskmelon, apple, grapes
Land 3	Orange, mungbean, coconut, papaya, pomegranate

through the lens of various research questions, providing a holistic view of its performance and potential.

5.1 Results Analysis of the Crop Filtration Phase

The proposed framework employs classification models to predict probabilities and regression models for yield prediction. The evaluation involves training and testing accuracy for the classification model and R^2, RMSE, and MAE for the regression models. Table 4 showcases the accuracies of different models in predicting the most suitable crop based on environmental conditions. The results as highlighted in the table, reveal that the proposed ANN model attains the highest training and testing accuracy values of 99.27% and 99.10%, respectively. In contrast, the decision tree records the lowest training and testing accuracies of 98.50% and 97.41%, respectively. Consequently, the proposed framework utilizes ANN for the crop filtration phase. Table 5 details the crops filtered by the ANN model for three lands, with flexibility for adjusting the number of crops based on user preference.

Figure 8a and b illustrate accuracy versus epoch and loss versus epoch graphs for the proposed ANN model, respectively. The loss curve indicates that the global optimal minima are attained at an early iteration stage. Moreover, the ANN-based crop recommendation model consistently produces more accurate results than conventional models. Table 6 provides a comprehensive comparison of the results achieved in the crop filtration phase against state-of-the-art work, affirming the superiority of the proposed model.

5.2 Results Analysis of the Yield Prediction Phase

Further, various regression models were applied to evaluate the second phase's performance, and the best-performing model was selected for yield prediction. Initially, data cleaning was applied to the crop yield dataset as different attributes have different measurements. The Min-Max Scaler was implemented from Python's Scikit-learn

Table 6 Comparative analysis (phase I) with state-of-the-art work

| Reference	Year	Crops	Features	Model	Result
Patel et al. [34]	2020	Mix crops	Soil	SVM, KNN	Acc.:93%
Jain et al. [35]	2020	10 crops	Rainfall, soil	RNN, RF	RMSE: 8.17%
Cheema et al. [15]	2021	10 crops	Soil	GSA	–
Patel et al. [36]	2021	Mix crops	Soil	ANN	Acc.:95.20%
Bakthavatchalam et al. [16]	2022	Mix crops	soil, weather	MLP	Acc.:98.22%
Mariammal et al. [19]	2021	9 crops	Soil, weather	Bagging	Acc.:95%
Ajoodha et al. [37]	2023	10 crops	Soil	RF	Acc.:91.1%
Gupta et al. [18]	2021	17 crops	Soil, humidity	K-means	Acc.:84%
Daniel et al. [23]	2022	Mix crops	Soil	SVM	Acc.:90%
Apat et al. [38]	2023	Mix crops	Soil, weather	Cat-boosting	Acc.:99%
Swathi et al. [26]	2023	6 crops	Soil	XGB, NB	AUC: 0.994, 0.993%
Bandi et al. [27]	2023	Mix crops	Soil	Ensemble	Acc.: 99.4, 0.993%
Shingade et al. [28]	2023	Mix crops	Soil, weather	RF	Acc.:95%
Proposed work (Phase I)	–	22 crops	Soil, weather	ANN	Acc.:99.10%
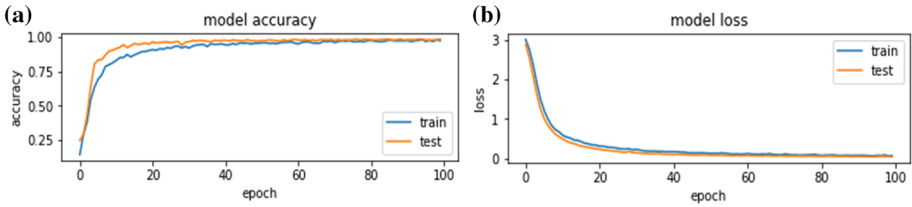
Fig. 8 ANN results **a** accuracy versus epoch **b** loss versus epoch

Table 7 Comparative analysis (phase II)

Metric/models	RF	DT	XGB	SVR	Ridge
R^2	**0.999**	0.998	**0.999**	0.883	0.998
MAE	**2.64e−4**	3.28e−4	2.97e−4	6.80e−4	3.01e−4
RMSE	**9.76e1**	1.17e2	**8.45e1**	2.95e2	1.45e2

library to obtain an accurate yield prediction. The dataset was rescaled using Eq. 5.

$$Y = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{5}$$

where Y is the rescaled value, X is the attributes value, X_{min} is the minimum value, and X_{max} is the maximum value of the attribute.

ML algorithms such as RF, DT, XGB, SVR, and ridge regression are applied to the preprocessed data for crop yield prediction. The performance analyses of these models are compared in Table 6, and the model is selected based on performance metrics, including MAE, RMSE, and R^2. The optimal scenario is when R^2 is maximized and MAE or RMSE is minimized. Further, comparisons of the achieved results are shown in Figs. 9 and 10. Figure 9a and b compare the models in terms of R^2 and MAE, respectively. The graphs shows that SVR is the worst performing model with the lowest R^2 and highest MAE values; other models have optimal values, whereas RF has the lowest MAE value. Similarly, Fig. 10 compares models in terms of RMSE. SVR has the highest RMSE value and the lowest model score. RF and XGB have the lowest RMSE values and with the highest model scores.

Tree-based models like RF, XGB, and DT exhibit more stable performance than others, owing to their ability to establish a stable and accurate decision boundary. Their robust performance is attributed to decision-making based on the outcomes of multiple trees, with majority voting contributing to precise predictions. RF has consistently demonstrated precision in various agricultural applications, particularly yield prediction, offering high accuracy, convenience, and practical utility in data modeling. It emerged as the best-performing model across multiple parameters and, thus, was selected for the yield prediction stage. Contrarily, SVR is the least effective model, as it is better suited for discrete problems for which it attempts to create the best-fit line using support vectors. Figures 9 and 10 visually highlight that RF delivers optimal performance with a 0.99 R^2 error. Consequently, Table 7 with highlighted
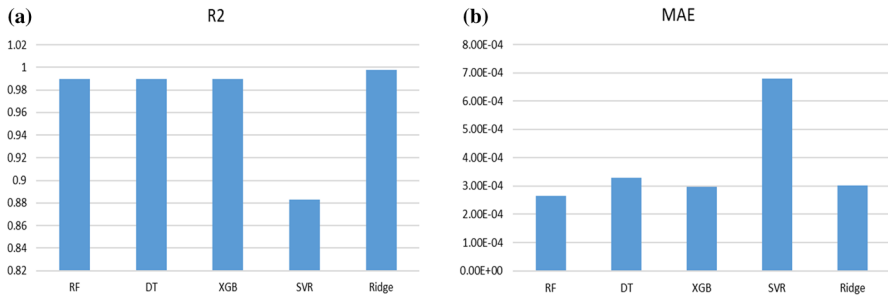
(a)

R2

(b)

MAE

Fig. 9 Performance comparison of various models in terms of **a** R^2 and **b** MAE
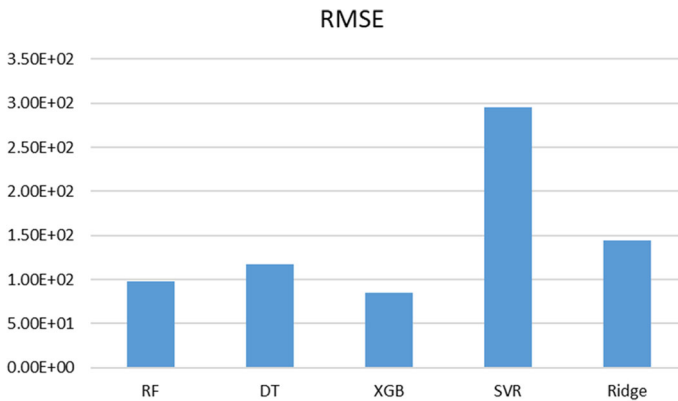
RMSE

Fig. 10 Performance comparison of various models in terms of RMSE

results, show that RF outperforms in the production of the crops obtained from the first phase.

5.3 Results Analysis of the Final Recommendations

The yield, cost of production, and market price of each filtered crop for Land 1 are shown in Table 8. The COP data were collected from a government website [39], and the MP data were collected from the agriculture commodity market [40]. It can be seen that of the five filtered crops, papaya, watermelon, and apple would be the most profitable for the framer. Hence, these crops are recommended to the farmer by the proposed framework as highlighted in the table. Further, Table 9 provides a comparative analysis of the results achieved in the yield prediction phase against state-of-the-art methods, affirming the efficiency of the proposed tree-based RF model.

5.4 Discussion

One of the noteworthy developments presented in this study is the formulation of a two-phase framework designed to recommend crops. In the initial phase, artificial neural

Table 8 Crop yield and profit

Crop	Cost of production (Rs. per ha)	Yield (kg/ha)	Market price (Rs. per kg)	Total profit (Rs.)
Watermelon	55,000	15,000	20	**2,45,000**
Papaya	2,10,000	27,000	20	**2,79,880**
Grapes	1,24,116	5500	60	2,05,884
Chickpea	11,250	800	60	36,750
Apple	1,50,000	6500	60	**2,40,000**

Table 9 Comparative analysis (phase II) with state-of-the-art work

References	Year	Features	Model	Crops	Result
Jain et al. [17]	2020	Soil, climate	SVM	Wheat, maize	Acc.: 85.62%, 75.64%
Khosla et al. [20]	2020	Rainfall	RNN, SVR	Bajra, maize rice, ragi	RMSE: 1709t, 1696t, 10,749t, 1731t
Gopal et al. [21]	2019	Area, Temp., etc	MLR-ANN	Paddy	R^2: 0.99
Elavarasan et al. [22]	2020	Soil, climate	RNN, Q-learning	Paddy	Acc. 93.7%
Iniyan et al. [24]	2022	Solar Rad., Temp., precipitation	Multilevel stacked ensemble regression	Corn, soyabean	MAE:6.63, 7.431 RMSE:10.545, 11.005
Kuradusenge et al. [25]	2023	Rainfall, Temp	RF	Potato, maize	RMSE:510.8, 129.9 R^2:0.875, 0.817
Shams et al. [29]	2024	Season, area, production	XAI-CROP	Mix crops	MAE:0.9874, MSE: 0.9412, R^2:0.94152
Olisha et al. [30]	2024	Soil, climate	DNNR	Corn	ARSE:0.0146t/ha, MAE:0.0209t/ha
Proposed work (Phase II)	–	Season, area, soil, weather, etc.	RF	17 crops	R^2: 0.99%, MAE: 2.64e−4, RMSE:9.7e1

network implementation helps filter crops based on their probabilities of utilizing farm-specific preferences. The random forest model also stands out for its exceptional performance in accurately suggesting crop yield. The results achieved by these models reflect the proposed framework's efficacy in situations requiring personalized farm recommendations. Furthermore, the following research questions (RQs) are framed to validate the applicability of SCR in current real-time farming settings.

- *RQ1: How does the smart crop recommendation framework utilizing an artificial neural network for crop filtration and random forest for yield prediction contribute to more accurate and informed crop recommendations compared to traditional farming practices?*

Traditional farming often relies on experience, intuition, and general practices

passed down through generations that might not align with current or specific on-site farm conditions. In contrast, the SCR framework is a data-driven decision-making tool that analyzes extensive datasets, incorporating weather patterns, soil health, historical yields, and market trends. Further, the ANN excels at recognizing intricate patterns within a dataset, enabling precise crop filtration based on a myriad of local input parameters. This allows for a nuanced understanding of the complex relationships between various factors, something traditional practices struggle with. As an ensemble learning method, random forest contributes to robust yield predictions by aggregating outputs from multiple decision trees. This enhances prediction accuracy and provides a reliable estimation of potential yields, surpassing the capabilities of traditional methods that often rely on intuition or experience. Hence, this data-driven approach enables more precise and personalized recommendations than other approaches.

- *RQ2: How does integrating local input parameters, such as weather, soil conditions, cost, and market prices, into the SCR framework affect the accuracy and relevance of crop recommendations for farmers?*
  By incorporating in-situ weather and soil condition data, the model adapts to the specific agro-climatic conditions of each farm. Considering costs and market prices ensures economic viability, helping farmers make informed decisions based on both yield potential and financial feasibility. This comprehensive integration not only refines the accuracy of crop predictions but also tailors recommendations to unique challenges and opportunities at the farm level, thus optimizing the overall effectiveness of the SCR framework.

- *RQ3: How does the proposed SCR framework address the challenges and limitations farmers face in traditional crop planning, and how does it contribute to sustainable agricultural practices?*
  The proposed framework addresses several challenges inherent to traditional crop planning, contributing significantly to sustainable agricultural practices. This heightened accuracy minimizes the risks associated with suboptimal crop choices, promoting resource efficiency and reducing financial losses for farmers. Furthermore, by incorporating sustainability metrics like soil features, the SCR framework encourages the cultivation of crops aligned with environmental conservation goals, optimizing resource utilization and minimizing ecological impact. In essence, the SCR framework enhances the economic viability of farming and promotes ecologically sustainable agricultural practices.

- *RQ4: How does the SCR framework align with current technological advancements and the need for modernization in the agricultural sector to meet the increasing demands for both quantity and quality of food?*
  The SCR framework utilizes machine learning models, including ANN and RF, to analyze extensive datasets and derive valuable insights in alignment with the broader trend of integrating technology into agriculture for improved productivity and decision-making. Additionally, by integrating local farm features, the framework embodies digital agriculture, adheres to modernization goals, implements personalized data-driven strategies, and addresses increasing food demands.

- *RQ5: What are the potential implications of implementing the proposed SCR framework on the agricultural sector's overall economic viability and produc-*

tivity, and how does it align with the long-term goals of ensuring food security and meeting the demands of a growing global population?

The implementation of the SCR framework holds profound implications for the overall economic viability and productivity of the agricultural sector. Below are several ways in which it aligns with long-term goals of ensuring food security and meeting the demands of a growing global population:

- *By cultivating crops with higher market demand and favorable growth conditions, farmers are better positioned to improve their income, contributing to the overall economic health of the agricultural sector.*
- *Precise recommendations tailored to specific agro-climatic conditions ensure that available resources are efficiently utilized, leading to higher yields per unit area and, consequently, improved overall productivity in the agricultural sector.*
- *By aligning crop recommendations with ecological considerations, the SCR framework contributes to sustainable farming practices crucial for the long-term health of the agricultural sector.*
- *Enhancing the precision of crop planning, mitigating risks associated with suboptimal choices, and adapting to changing conditions contribute to a more resilient and secure food supply chain.*
- *With its data-driven and technologically adaptable approach, the SCR framework aligns with global efforts to modernize agriculture.*

These research directions validate the practicality of the proposed framework, especially in contexts characterized by resource limitations and constraints. Beyond its basic architecture, the framework fulfills a dual function of crop selection and yield prediction, eliminating the need for separate applications for these tasks. This streamlined approach enhances the model's efficiency and practicality.

6 Conclusion

India stands as a leading producer of agricultural goods, yet there exists untapped potential for optimizing productivity. If crop yield and return are to be elevated, it is imperative to pinpoint factors that can enhance the current agricultural landscape. A critical determinant in crop production is the selection of the most suitable crop based on geographical and geological conditions. However, there is a noticeable deficiency in scientific agricultural literacy within the farming community, leading to reliance on conventional practices. Addressing such challenges through computational means is of paramount importance. Machine learning algorithms have emerged as a transformative bridge, reducing the knowledge gap between agricultural experts and farmers.

This research proposes a crop recommendation system that can identify the most fitting crops for specific regional conditions. The envisioned solution relies on a standardized dataset acting as a domain expert to inform decisions. Subsequently, a computational process harnesses this agricultural dataset, employing machine learning techniques to construct trained models. The proposed smart crop recommendation framework, grounded in machine learning, empowers farmers with well-informed

decisions on optimal crop selection. Uniquely positioned to handle challenges at the farm level, the two-phase SCR framework comprehensively analyzes local factors, recognizing the dynamic nature of diverse agricultural features. In the initial phase, the framework achieved an impressive accuracy of 99.10% using artificial neural networks. Additionally, in the next phase, the random forest demonstrated high performance, with a 0.99 R^2 error metric. The experimental results attest to the proposed framework's efficacy, positioning it as a practical and efficient real-time recommendation solution. In terms of practical implications, the simple and lightweight design of the suggested framework offers potential for future integration with handheld devices. Moreover, there are opportunities for future explorations of the expansion of the application's utility (e.g., incorporating a closed-loop supply chain, predicting fertilizer needs, and recognizing plant diseases), thus providing more comprehensive solutions for crop and soil management.

Data Availability The dataset used is publicly available as cited in the article.

Code Availability Code will be made available on request.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Wikipedia: agriculture in India. https://en.wikipedia.org/wiki/Agriculture_in_India. Accessed 16 Jan 2024
2. Minister of Agriculture & Farmers Welfare (2023) Contribution of agricultural sector in GDP. https://www.pib.gov.in/PressReleasePage.aspx?PRID=1909213. Accessed 16 Jan 2024
3. Food and Agriculture Organisation of the United Nations. India at a glance. https://www.fao.org/india/fao-in-india/india-at-a-glance/en/. Accessed 16 Jan 2024
4. Olson DL, Shi Y, Shi Y (2007) Introduction to business data mining, vol 10
5. Bhola A, Kumar P (2023) Performance evaluation of different machine learning models in crop selection. In: Robotics, control and computer vision, pp 207–217. https://doi.org/10.1007/978-981-99-0236-1_16
6. Verma S, Kumar P, Singh JP (2023) A unified lightweight CNN-based model for disease detection and identification in corn, rice, and wheat. IETE J Res. https://doi.org/10.1080/03772063.2023.2181229
7. Bhola A, Verma S, Kumar P (2023) A comparative analysis of deep learning models for cucumber disease classification using transfer learning. J Curr Sci Technol 13(1):23–35. https://doi.org/10.14456/jcst.2023.3
8. Shi Y (2022) Advances in big data analytics. Adv Big Data Anal. https://doi.org/10.1007/978-981-16-3607-3
9. Tien JM (2017) Internet of things, real-time decision making, and artificial intelligence. Ann Data Sci 4:149–178. https://doi.org/10.1007/s40745-017-0112-5
10. Sinha A, Shrivastava G, Kumar P (2019) Architecting user-centric internet of things for smart agriculture. Sustain Comput Inform Syst 23:88–102. https://doi.org/10.1016/j.suscom.2019.07.001

11. Kashyap PK, Kumar S, Jaiswal A, Prasad M, Gandomi AH (2021) Towards precision agriculture: Iot-enabled intelligent irrigation systems using deep learning neural network. IEEE Sens J 21(16):17479–17491. https://doi.org/10.1109/JSEN.2021.3069266

12. Wikipedia: digital agriculture. https://en.wikipedia.org/wiki/Digital_agriculture. Accessed 16 Jan 2024

13. Shi Y, Tian Y, Kou G, Peng Y, Li J (2011) Optimization based data mining: theory and applications. https://doi.org/10.1007/978-0-85729-504-0

14. Latino ME, Menegoli M, Corallo A (2022) Agriculture digitalization: a global examination based on bibliometric analysis. IEEE Trans Eng Manag. https://doi.org/10.1109/TEM.2022.3154841

15. Cheema SS, Singh A, Gritli H (2021) Optimal crop selection using gravitational search algorithm. Math Probl Eng. https://doi.org/10.1155/2021/5549992

16. Bakthavatchalam K, Karthik B, Thiruvengadam V, Muthal S, Jose D, Kotecha K, Varadarajan V (2022) Iot framework for measurement and precision agriculture: predicting the crop using machine learning algorithms. Technologies 10(1):13. https://doi.org/10.3390/technologies10010013

17. Jain K, Choudhary N (2022) Comparative analysis of machine learning techniques for predicting production capability of crop yield. Int J Syst Assur Eng Manag 13(1):583–593. https://doi.org/10.1007/s13198-021-01543-8

18. Gupta R, Sharma AK, Garg O, Modi K, Kasim S, Baharum Z, Mahdin H, Mostafa SA (2021) Wb-cpi: weather based crop prediction in India using big data analytics. IEEE Access 9:137869–137885. https://doi.org/10.1109/ACCESS.2021.3117247

19. Mariammal G, Suruliandi A, Raja S, Poongothai E (2021) Prediction of land suitability for crop cultivation based on soil and environmental characteristics using modified recursive feature elimination technique with various classifiers. IEEE Trans Comput Soc Syst 8(5):1132–1142. https://doi.org/10.1109/TCSS.2021.3074534

20. Khosla E, Dharavath R, Priya R (2020) Crop yield prediction using aggregated rainfall-based modular artificial neural networks and support vector regression. Environ Dev Sustain 22(6):5687–5708. https://doi.org/10.1007/s10668-019-00445-x

21. Gopal PM, Bhargavi R (2019) A novel approach for efficient crop yield prediction. Comput Electron Agric 165:104968. https://doi.org/10.1016/j.compag.2019.104968

22. Elavarasan D, Vincent PD (2020) Crop yield prediction using deep reinforcement learning model for sustainable agrarian applications. IEEE Access 8:86886–86901. https://doi.org/10.1109/ACCESS.2020.2992480

23. Daniel J, Shyamala R, Pugalenthi R, Mohan Kumar P (2022) Ranc-crop recommendation attributed to soil nutrients and stock analysis using machine learning. IETE J Res. https://doi.org/10.1080/03772063.2022.2060868

24. Iniyan S, Jebakumar R (2022) Mutual information feature selection (MIFS) based crop yield prediction on corn and soybean crops using multilayer stacked ensemble regression (mser). Wirel Pers Commun 126(3):1935–1964. https://doi.org/10.1007/s11277-021-08712-9

25. Kuradusenge M, Hitimana E, Hanyurwimfura D, Rukundo P, Mtonga K, Mukasine A, Uwitonze C, Ngabonziza J, Uwamahoro A (2023) Crop yield prediction using machine learning models: case of Irish potato and maize. Agriculture 13(1):225. https://doi.org/10.3390/agriculture13010225

26. Swathi T, Sudha S (2023) Crop classification and prediction based on soil nutrition using machine learning methods. Int J Inf Technol 15(6):2951–2960. https://doi.org/10.1007/s41870-023-01345-0

27. Bandi R, Likhit MSS, Reddy SR, Bodla SR, Venkat VS (2023) Voting classifier-based crop recommendation. SN Comput Sci 4(5):516. https://doi.org/10.1007/s42979-023-01995-8

28. Shingade SD, Mudhalwadkar RP (2023) Sensor information-based crop recommendation system using machine learning for the fertile regions of Maharashtra. Concurr Comput Pract Exp 35:7774

29. Shams MY, Gamel SA, Talaat FM (2024) Enhancing crop recommendation systems with explainable artificial intelligence: a study on agricultural decision-making. Neural Comput Appl. https://doi.org/10.1007/s00521-023-09391-2

30. Olisah C, Smith L, Smith M, Morolake L, Ojukwu O (2024) Corn yield prediction model with deep neural networks for smallholder farmer decision support system. arXiv preprint arXiv:2401.03768. https://doi.org/10.48550/arXiv.2401.03768

31. Devi M, Malik D, Mehala V, Mishra P (2020) Measuring variability and factors affecting the agricultural production: a ridge regression approach. Ann Data Sci. https://doi.org/10.1007/s40745-020-00274-0

32. Ingle A. Crop recommendation dataset. https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset. Accessed 03 Jan 2023

33. Minister of Agriculture & Farmers Welfare. Area production and yield information system. https://aps.dac.gov.in/Home.aspx?ReturnUrl=%2f. Accessed 16 Jan 2024

34. Patel K, Patel HB (2020) A state-of-the-art survey on recommendation system and prospective extensions. Comput Electron Agric 178:105779. https://doi.org/10.1016/j.compag.2020.105779

35. Jain S, Ramesh D (2020) Machine learning convergence for weather based crop selection. In: 2020 IEEE international students' conference on electrical, electronics and computer science (SCEECS). IEEE, pp 1–6. https://doi.org/10.1109/SCEECS48394.2020.75

36. Patel K, Patel HB (2021) A comparative analysis of supervised machine learning algorithm for agriculture crop prediction. In: 2021 fourth international conference on electrical, computer and communication technologies (ICECCT). IEEE, pp 1–5. https://doi.org/10.1109/ICECCT52121.2021.9616731

37. Ajoodha R, Mufamadi TO (2023) Crop recommendation using machine learning algorithms and soil attributes data. In: Proceedings of 3rd international conference on artificial intelligence: advances and applications: ICAIAA 2022, pp 31–41

38. Apat SK, Mishra J, Raju KS, Padhy N (2023) An artificial intelligence-based crop recommendation system using machine learning. J Sci Ind Res (JSIR) 82(05):558–567. https://doi.org/10.56042/jsir.v82i05.1092

39. Cabinet Committee on Economic Affairs (CCEA) (2022) CCEA approves minimum support prices (MSP) for Kharif Crops for marketing season 2022–23. https://pib.gov.in/PressReleasePage.aspx?PRID=1832172. Accessed 16 Jan 2024

40. Agriculture marketing. https://agmarknet.gov.in/PriceAndArrivals/CommodityDailyStateWise.aspx. Accessed 03 Jan 2023